A comparison between Numerical Weather Prediction and Machine Learning-based forecasts in the Arctic

Stockholm University
Department of Meteorology

Astrid Björndal with supervisors Gunilla Svensson & Linus Magnusson
astridmmbjorndal@gmail.com

Abstract

Increasing activity in the Arctic calls for improved forecast quality and understanding of weather phenomena. Since the 1950's computational Numerical Weather Prediction (NWP) have been a great tool for research, the general population and societal functions. Recently a new technology has been introduced into forecasting meteorology: Machine Learning (ML). A continuous and rapid development of ML forecasting models is executed by many prominent technology companies; Google DeepMind has developed GraphCast, Nvidia has FourCastNet and Huawei has PanguWeather. The European Centre for Medium-range Weather Forecasts (ECMWF) has developed a ML version of their Integrated Forecasting system (IFS), called AIFS. All listed ML models are trained on ERA5 Reanalysis with a 0.25° resolution. In the Arctic, model performance needs more evaluation, in this report, the forecast temperature, geopotential height, specific humidity and wind speed at different atmospheric levels are evaluated against analysis and observations. The time period and evaluation area was selected to be colocated with the ARTofMELT Arctic expedition (Tjernström et al., 2024). The analysis comparison integrated a closed region from 80°N to 88°N (60°Wand 60°E), the mean RMSE is lower for IFS the first three days of forecast, and ML performs better with increased forecast times. The mean bias of IFS was positive for temperature and geopotential height, while ML models had more negative and varying bias with longer forecast times. Inter-variable correlation in ML models were compared to physics-based IFS, and less connection was found between variables at different atmospheric levels in ML models. For further research, the assimilation of satellite data could be improved using new techniques, as well as increased resolution for ML-based forecasts.

Background

Predicting weather is a challenge, specifically the Arctic weather can be difficult; even more so in a changing climate. It is predicted that activities, like shipping and tourism, closer to the North Pole will increase in the future (Jung et al., 2016). This needs high quality weather forecasts for the best use. Observations are crucial for improving forecast quality, especially the Arctic needs better integration into numerical models. Since the 1950s, when the first computer-created weather forecast was made, Numerical Weather Prediction (NWP) has developed from physics to a whole new field of science. Using machine learning (ML) to make weather forecast has during recent years been developed constantly and rapidly (Ben Bouallègue et al., 2024). There are several ML forecasts available and easily accessible to the public. Both traditional NWP and MLWP work auto-regressively and uses model analysis data to compute the next time step.

Research question

- How well does Numerical Weather Prediction and machine learning-based forecasts perform in the Arctic during the ARTofMELT campaign?

How is it done?

- The quality of NWP model IFS and MLWP models *PanguWeather*, *FourCastNet*, *GraphCast* and *AIFS* were evaluated and compared against observations from Arctic expedition: ARTofMELT (Tjernström et al., 2024).

Forecast models

	Company	Model	Label
MLWP	Huawei	PanguWeather	pguw
MLWP	Nvidia	FourCastNet	sfno
MLWP	Google DeepMind	GraphCast	dmgc
MLWP	ECMWF	AIFS	aifs
NWP	ECMWF	IFS	0001

Theory & Methods

NWP – Integrated Forecasting System IFS

- European Centre for Medium-range Weather Forecasts ECMWF
- Available in 0.1° grid resolution (0,25° grid used here for all)
- Data assimilation of analysis step
- Model physics using numerical equations of state

Machine Learning-based forecasts

- No numerical equations solved, trained on ERA5 data in previous climate
- Initial analysis and uses data-driven models
- Model run-time is 10 000 times faster than IFS

$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(F_i - O_i)^2}$ Root Mean Square Error verified against analysis, comparing performance for the forecast lead time and different weather parameters.

Parameter		Unit
2 m temperature	2t0	°C
Wind speed 10 m above surface	ws10	m/s
850hPa temperature	t850	°C
850hPa specific humidity	q850	g/kg
500 hPa geopotential height	z500	m

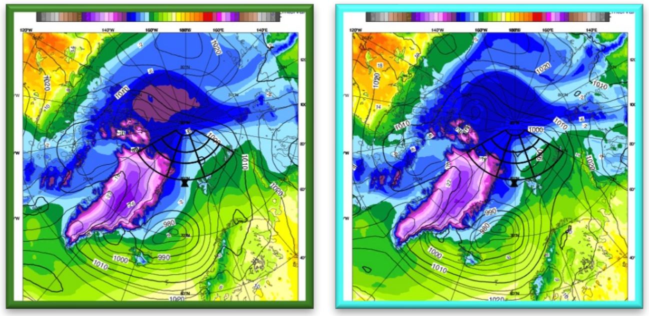

Figure 1: Forecast of 850hPa temperature in MLWP GraphCast (left) and NWP IFS (right).

Results

Mean of RMSE (Figure 2) :

- IFS in general integrated the analysis better, however with longer lead times it performs worse than MLWP.
- The overall highest performing MLWP are *AIFS* and *GraphCast*.

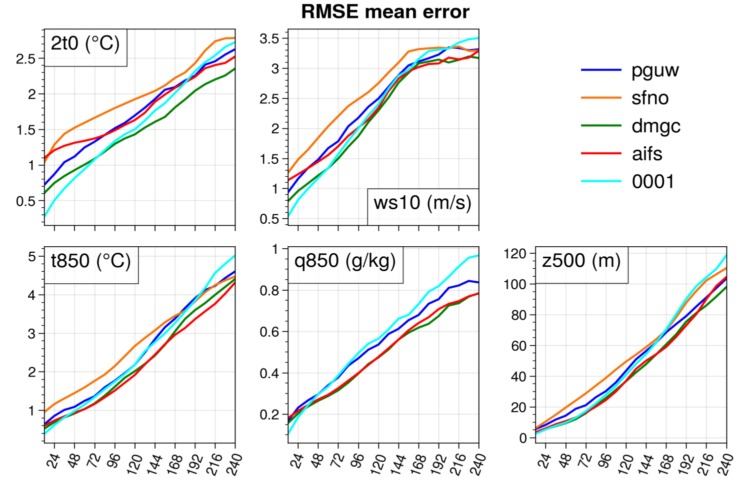

Figure 2: M ean of RMSE for the forecast time steps for all models over the whole period between 28th of April and 18th of June.

Magnitude of RMSE compared to observations of wind from ARTofMELT:

- Storm case on doy 146 was not well captured by any model 72h in advance.
- Some false high wind speeds around doy 140.

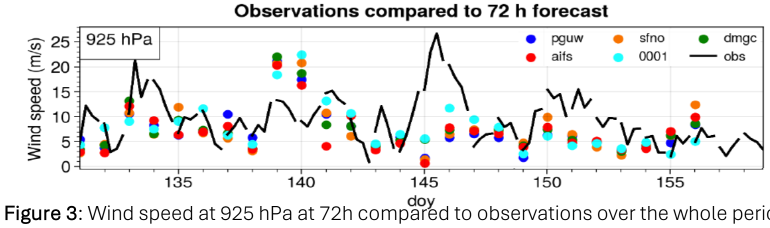

Figure 3: Wind speed at 925 hPa at 72h compared to observations over the whole period between 28th of April and 18th of June.

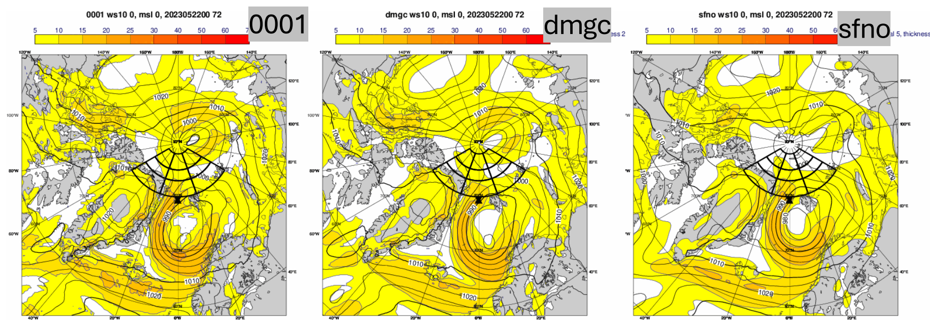

Figure 4: Wind speed at 10 m above surface at 72h for IFS (left), *GraphCast* (middle) and *FourCastNet* (right).

Discussion

- Synoptically the polar region is covered more accurately (Fig. 4) with more detail in the different investigated parameters.
  - Different grid solutions.
- IFS has a more stable linear increase of mean RMSE the first 7 days compared to MLWP forecast models
  - Generally, less extreme errors for IFS → numerically contained
- Many case studies find similar patterns for the model performance
  - Not done in the Arctic, all models need improvements in the Arctic
  - IFS has known issues with e.g. temperature in the Arctic region, MLWP trained on ERA5 data (based on IFS) so these errors can propagate in the ML training phase.
- The grid solution affects the synoptic structure, however the integrated error of the analysed region show errors of same magnitude.

Summary and Conclusion

Combining synoptic scale analysis and error analysis gives a trend of all investigated MLWP perform in a comparable way to the conventional NWP model IFS for both shorter and longer forecasts for the ARTofMELT campaign.
Especially, GraphCast was the model scoring most like IFS in both the statistical and synoptic evaluation against analysis. There are room for improvements since biases were found among other errors and the presentation of Arctic climate. Further research and development is crucial for predicting all variables in a satisfactory way through all atmospheric levels, around the whole globe, with high resolution and run time.

References

Ben Bouallègue, Z., Clare, M. C., Magnusson, L., Gascon, E., Maier-Gerber, M., Janoušek, M., Rodwell, M., Pinault, F., Dramsch, J. S., Lang, S. T., et al. (2024). The rise of data-driven weather orecasting: A first statistical assessment of machine learning-based weather forecasts in an operational-like context. Bulletin of the American Meteorological Society.

Jung, T., Gordon, N. D., Bauer, P., Bromwich, D. H., Chevallier, M., Day, J. J., Dawson, J., Doblas-Reyes, F., Fairall, C., Goessling, H. F., et al. (2016). Advancing polar prediction capabilities on daily to seasonal time scales. Bulletin of the American Meteorological Society, 97 (9), 1631–1647.

Tjernström, M., Zieger, P., & Murto, S. (2024). Arctic spring and the onset of. sea-ice melt: Early impressions from the artofmelt expedition. EGU24, (EGU24-15627).